# Machine Learning
# Homework Assignment 2

Nishika Chopra nc2259 N18598308
Karanpreet Singh Wadhwa ksw352 N13337853

# Machine Learning
## Homework -2
### PART 1

Ans 1) (a) Given :-

$$\text{True Positive Rate} = \frac{\#\text{ True Positive}}{\#\text{ whose correct label is } +}$$

$$= \frac{\#TP}{\#TP + \#FN}$$

According to the confusion matrix

$TP = 0$ and $FN = 3$

$$\text{True Positive Rate} = \frac{0}{0+3}$$

$$= 0 \quad \underline{\text{Ans.}}$$

(b) Given :-

$$\text{False Positive Rate} = \frac{\#\text{ False Positive}}{\#\text{ whose correct label is } -}$$

$$= \frac{\#FP}{\#FP + \#TN}$$

According to the confusion matrix

$FP = 1$ $TN = 2$

$$\text{False Positive Rate} = \frac{1}{1+2} = \frac{1}{3}$$

$\underline{\text{Ans.}}$

Scanned by CamScanner

(C) Accuracy = $\frac{\text{\# examples predicted correct}}{\text{\# of total predictions}}$

$$= \frac{TP + TN}{TP + FP + TN + FN}$$

Given,

$TP = 56$ $FP = 1$

$TN = 41$ $FN = 2$

Hence, Accuracy $= \frac{56 + 41}{56 + 1 + 41 + 2}$

$$= \frac{97}{100}$$

$= \cancel{0.95}\ 0.97$

or 97 % Ans

ans2) Given :-

$g_1(x_1, x_2) = 5x_2 + 3x_1 - 4 \equiv C_1$

$g_2(x_1, x_2) = -3x_2 + 2x_1 - 6 \equiv C_2$

when there are two classes, we define a single discriminant

$g(x) = g_1(x) - g_2(x)$

and we

~~to~~ choose $\begin{cases} C_1 & \text{if } g(x) > 0 \\ C_2 & \text{otherwise} \end{cases}$

In this question, we will define

$$g(x_1, x_2) = g_1(x_1, x_2) - g_2(x_1, x_2)$$

$$= (5x_2 + 3x_1 - 4) - (-3x_2 + 2x_1 - 6)$$

$$= (5x_2 + 3x_2) + (3x_1 - 2x_1) + (-4 + 6)$$

$$= 8x_2 + x_1 + 2$$

Now, to prove this ~~is the correct~~ new discriminant function will result in correct assignment we take 2 examples

~~Example 1: -~~

~~Let $x_1 = 5x_2 = 5$~~

~~$x_1 = 2$ and~~

~~$g_1(x) =$~~

## Example 1:

Let $x_2 = 2$ and $x_1 = 5$

$$g_1(x_1, x_2) = 5x_2 + 3x_1 - 4$$
$$= 5 \times 2 + 3 \times 5 - 4$$
$$= 10 + 15 - 4$$
$$= 21$$

$$g_2(x_1, x_2) = -3x_2 + 2x_1 - 6$$
$$= -3 \times 2 + 2 \times 5 - 6$$
$$= \cancel{-6 + 10 - 6} = -6 + 10 - 6$$
$$= -2$$

$$g(x_1, x_2) = 8x_2 + x_1 + 2$$
$$= 8 \times 2 + 5 + 2$$
$$= 16 + 5 + 2$$
$$= 23$$

Thus,
Since $g_1(x_1, x_2) > g_2(x_1, x_2)$
and $g(x_1, x_2) > 0$

Both the cases will assign this to the same class that is $C_1$.

Example 2:

Let $x_2 = -2$ and $x_1 = 1$

$$g_1(x_1, x_2) = 5x_2 + 3x_1 - 4$$
$$= 5 \times (-2) + 3 \times 1 - 4$$
$$= -10 + 3 - 4$$
$$= -11$$

$$g_2(x_1, x_2) = -3x_2 + 2x_1 - 6$$
$$= (-3)(-2) + 2 \times 1 - 6$$
$$= 6 + 2 - 6$$
$$= 2$$

$$g(x_1, x_2) = 8x_2 + x_1 + 2$$
$$= 8(-2) + 1 + 2$$
$$= -16 + 1 + 2$$
$$= -13$$

Therefore, $g_1(x_1, x_2) < g_2(x_1, x_2)$

and $g(x_1, x_2) < 0$

Hence again both the cases assigned the example to the same class that is $C_2$

Hence Proved.

Ans.

ans3) (a) Given

Cost of false negative $= a$
$= 5$

Cost of false positive $= b$
$= 2$

$$P(-) = a * P(c=+|x)$$
$$P(+) = b * P(C=-|x)$$

~~P(spa~~ Let spam $= +$
not spam $= -$

$$P(+|x_1, x_2) = \frac{1}{1+e^{-(3x_2 - 2x_1+1)}}$$
$$= \frac{1}{1+e^{-(3\times 2 - 2\times 3 + 1)}}$$

$\{\because x_1 = 3$ and $x_2 = 2\}$

$$P(+|x_1, x_2) = \frac{1}{1+e^{-1}}$$
$$= \frac{1}{1+0.36788}$$
$$= 0.731$$

$$P(-|x_1, x_2) = 1 - P(+|x_1, x_2)$$
$$= 1 - 0.731$$
$$= 0.269$$

$$P(-) = a * P(+|x_1, x_2)$$
$$= 5 \times 0.731$$
$$= 3.655$$

$$P(+) = b * P(-|x_1, x_2)$$
$$= 2 \times 0.269$$
$$= 0.538$$

Hence P(+) that is predicting + that is spam has a smaller expected risk.

4(a) Bias of an estimator $d = \frac{\sum_{x \in X} x}{N+1}$, $\theta = \mu$

$$b_\theta = E(d) - \theta$$

$$b_\theta = E\left(\frac{\sum_{x \in X} x}{N+1}\right) - \theta$$

$$b_\theta = \frac{\sum_{x \in X} E(x)}{N+1} - \theta$$

$$= \frac{\sum_{x \in X} \mu}{N+1} - \theta$$

$$= \frac{N\mu}{N+1} - \mu = \frac{N\mu}{N+1} - \frac{(N+1)\mu}{N+1}$$

$$= \frac{\cancel{N\mu} - \cancel{N\mu} - \mu}{N+1} = \frac{-\mu}{N+1}$$

Scanned by CamScanner

b) No the ~~bias~~ [Answer] would Not change g. In this just the ^ symbol to show mean would change.

$\theta$ for exponential distribution $= \frac{1}{\lambda} = \beta$

$$b_\theta = E(d) - \theta$$

$$= \frac{E\left(\sum_{x \in X} x\right)}{N+1} - \beta$$

$$= \frac{N\beta}{N+1} - \beta$$

$$= \frac{\cancel{N\beta} - \cancel{N\beta} - \beta}{N+1}$$

$$= -\frac{\beta}{N+1}$$

ans 5) Dataset

| $x_1$ | $x_2$ | Label |
|---|---|---|
| 2.7 | 4.8 | + |
| 3.2 | 5.1 | + |
| -0.4 | -0.3 | + |
| 0.6 | 0.5 | - |
| 1.8 | 2.8 | - |
| 2.1 | 4.31 | - |

(a) To implement the multivariate Gaussian we divide the data into two datasets, One for +ve class and other for -ve class.

| $x_1$ | $x_2$ | label |
|---|---|---|
| 2.7 | 4.8 | + |
| 3.2 | 5.1 | + |
| -0.4 | -0.2 | + |

| $x_1$ | $x_2$ | label |
|---|---|---|
| 0.6 | 0.5 | - |
| 1.8 | 2.8 | - |
| 2.1 | 4.3 | - |

$$\mu_+ = \begin{bmatrix} (2.7+3.2-0.4)/3 \\ (4.8+5.1-0.2)/3 \end{bmatrix} \qquad \mu_- = \begin{bmatrix} (0.6+1.8+2.1)/3 \\ (0.5+2.8+4.3)/3 \end{bmatrix}$$

$$\mu_+ = \begin{bmatrix} 1.83 \\ 3.23 \end{bmatrix} \qquad \mu_- = \begin{bmatrix} 1.5 \\ 2.53 \end{bmatrix}$$

Now we calculate covariance using $\mu$ for each dataset.

$$\Sigma_+ = \sum_{t=1}^{N} \frac{(x^t - \mu_+)(x^t - \mu_+)^T}{N}$$

$$= \frac{1}{3}\left[ \begin{bmatrix}0.87\\1.57\end{bmatrix}\begin{bmatrix}0.87 & 1.57\end{bmatrix} + \begin{bmatrix}1.37\\1.87\end{bmatrix}\begin{bmatrix}1.37 & 1.87\end{bmatrix} + \begin{bmatrix}-2.23\\-3.43\end{bmatrix}\begin{bmatrix}-2.23 & -3.43\end{bmatrix} \right]$$

$$= \frac{1}{3}\begin{bmatrix}7.6 & 11.6\\ 11.6 & 17.7\end{bmatrix}$$

$$= \begin{bmatrix}2.54 & 3.86\\ 3.86 & 5.9\end{bmatrix}$$

$$\Sigma_- = \sum_{t=1}^{N} \frac{(x^t - \mu_-)(x^t - \mu_-)^T}{N}$$

$$= \frac{1}{3}\left[ \begin{bmatrix}-0.9\\-2.03\end{bmatrix}\begin{bmatrix}0.9 & -2.03\end{bmatrix} + \begin{bmatrix}0.3\\0.27\end{bmatrix}\begin{bmatrix}0.3 & 0.27\end{bmatrix} + \begin{bmatrix}0.6\\1.77\end{bmatrix}\begin{bmatrix}0.6 & 1.77\end{bmatrix} \right]$$

$$= \frac{1}{3}\begin{bmatrix}1.26 & 2.97\\ 2.97 & 7.3267\end{bmatrix}$$

$$= \begin{bmatrix}0.42 & 0.99\\ 0.99 & 2.44\end{bmatrix}$$

Ans.

(b) The pdf for d dimensional multivariate

$$\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$\mu$ = mean

$\Sigma$ = covariance

$x$ = ~~m~~ new data

$$X = \begin{bmatrix} 1.6 \\ 2.3 \end{bmatrix}$$

we calculate

$$(x-\mu) = \begin{bmatrix} \cancel{1.6 - 1.83} \end{bmatrix} \begin{bmatrix} 1.6 - 1.83 \\ 2.3 - 3.2 \end{bmatrix} = \begin{bmatrix} -0.23 \\ -0.9 \end{bmatrix}$$

$$\Sigma^{-1} = \frac{1}{0.0893}\begin{bmatrix} 6.14 & -3.93 \\ -3.93 & 2.53 \end{bmatrix}$$

$$\log p(x|i) = \log\left(\frac{1}{2\pi|\Sigma|^{1/2}}\right) - \frac{1}{2}\begin{bmatrix} -0.23 & -0.9 \end{bmatrix}\begin{bmatrix} 6.14 & -3.93 \\ -3.93 & 2.53 \end{bmatrix}\frac{1}{0.0893}\begin{bmatrix} -0.23 \\ 0.9 \end{bmatrix}$$

$$= -0.27 - \frac{1}{2\times 0.0893}\begin{bmatrix} 2.1248 & -1.3731 \end{bmatrix}\begin{bmatrix} -0.23 \\ 0.9 \end{bmatrix}$$

$$= -0.27 - \frac{0.747086}{0.1786}$$

$$= -0.27 - 4.183$$

$$= -4.453$$

Ans-

$$\log(pdf) = \log\frac{1}{2\pi(\Sigma)^{\frac{1}{2}}} - \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$$

For -ve

$$= \log\frac{1}{2\pi(0.211)} - \frac{1}{2}\begin{bmatrix}0.1 & -0.23\end{bmatrix}\begin{bmatrix}2.44 & -0.99 \\ -0.99 & 0.42\end{bmatrix}\begin{bmatrix}0.1 \\ -0.23\end{bmatrix}$$

$$(x-\mu)^T = \begin{bmatrix}1.6-1.5 \\ 2.3-2.53\end{bmatrix} = \begin{bmatrix}0.1 \\ 0.23\end{bmatrix} = \begin{bmatrix}0.1 & -0.23\end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix}0.1 \\ -0.23\end{bmatrix}$$

$$= -0.1224 - \frac{1}{0.0894}\begin{bmatrix}0.4717 & 0.9756\end{bmatrix}\begin{bmatrix}0.1 \\ -0.23\end{bmatrix}$$

$$= -0.1224 + \frac{0.092158}{0.0894} = 0.90$$

$\log P(x|+) = -4.453$

$\log p(x|-) = 0.9$

As $p(x|-) > p(x|+)$

$x$ belongs to -ve.

Scanned by CamScanner

ans 6)

| Height $(x^t)$ | Stories $(r^t)$ | $(x^t)^2$ | $x^t r^t$ |
|---|---|---|---|
| 1050 | 57 | 1102500 | 59850 |
| 428 | 28 | 183184 | 11984 |
| 362 | 26 | 131044 | 9412 |
| 529 | 40 | 279841 | 21160 |
| 790 | 60 | 624100 | 47400 |
| 401 | 22 | 160801 | 8822 |
| 380 | 38 | 144400 | 14440 |
| 1454 | 110 | 2114116 | 159940 |
| 1127 | 100 | 1270129 | 112700 |
| 700 | 46 | 490000 | 32200 |

$\Sigma(x^t) = 7221$

$\Sigma(x^t)^2 = 6500115$

$\Sigma(r^t) = 527$

$\Sigma(x^t r^t) = 477908$

$N = 10$

$$A = \begin{bmatrix} N & \Sigma x^t \\ \Sigma x^t & \Sigma (x^t)^2 \end{bmatrix} \qquad \omega = \begin{bmatrix} \omega_0 \\ \omega_1 \end{bmatrix}$$

$$Y = \begin{bmatrix} \Sigma r^t \\ \Sigma r^t x^t \end{bmatrix}$$

$\therefore W = A^{-1} Y$

Scanned by CamScanner

$$W = \begin{bmatrix} 10 & 7221 \\ 7221 & 6500115 \end{bmatrix}^{-1} \begin{bmatrix} 527 \\ 477908 \end{bmatrix}$$

$$= \begin{bmatrix} -1.976 \\ 0.0757 \end{bmatrix}$$

$w_0 = -1.976$ $\quad$ $w_1 = 0.0757$

(a) $g(x) = 0.0757x - 1.976$

Ans.

(b) Now,

when $x = 475$

$g(x) = 34$

Thus when the height is 475 then the predicted stories comes out to be 34

Ans.

ans7) (a) Original data | Scaled data

| $x_1$ | $x_2$ | Label | $x_1$ | $x_2$ | label |
|---|---|---|---|---|---|
| 2.5 | 42 | + | 0.68 | 0.827 | + |
| 3.8 | 51 | + | 1.00 | 1.00 | + |
| −0.3 | −1 | + | 0.00 | 0.00 | + |
| 0.7 | 3 | − | 0.24 | 0.077 | − |
| 1.6 | 26 | − | 0.46 | 0.519 | − |
| 2.3 | 41 | − | 0.63 | 0.808 | − |

ans.

Scaling the second using

$$\frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}$$

~~for~~ for $x_2$:

~~min~~ $x_2^{min} = -1$

$x_2^{max} = 51$

Hence after scaling we get

(b) ~~$x_1 =$~~ $x = \begin{bmatrix} 3.9 \\ 4 \end{bmatrix}$

$$\frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}$$

So,

for $x_1$

$x_1^{min} = -0.3$

$x_1^{max} = 3.8$

for $x_2$

$x_2^{min} = -1$

$x_2^{max} = 51$

Scaled values $\Rightarrow$

$$x' = \begin{bmatrix} 1.024 \\ 0.096 \end{bmatrix}$$

E.d. wt. eg. 1 = $\sqrt{(1.024-0.68)^2 + (0.096-0.827)^2} = 0.8078$

2 = 0.904219

3 = 1.02849

4 = 0.78423

5 = 0.70554

6 = 0.8154

As the example is closer to 5th sample, the final answer is -ve